Title: Robust clustering of depression subtypes – Linking symptoms and morphology

Authors: L Jollans, P Sämann, T Brückl, N Rost, WE Craighead, KS Choi, HS Mayberg, BW Dunlop & E Binder

Introduction: Major depressive disorder (MDD) is the most common psychiatric disorder, and is one of the leading causes of years lived with disability [1]. Identification of treatments most likely to succeed with each individual MDD patient remains challenging. Further subdivision of MDD patients into clinically and/or biologically distinct groups would provide a valuable avenue for future improvement in treatment targeting.

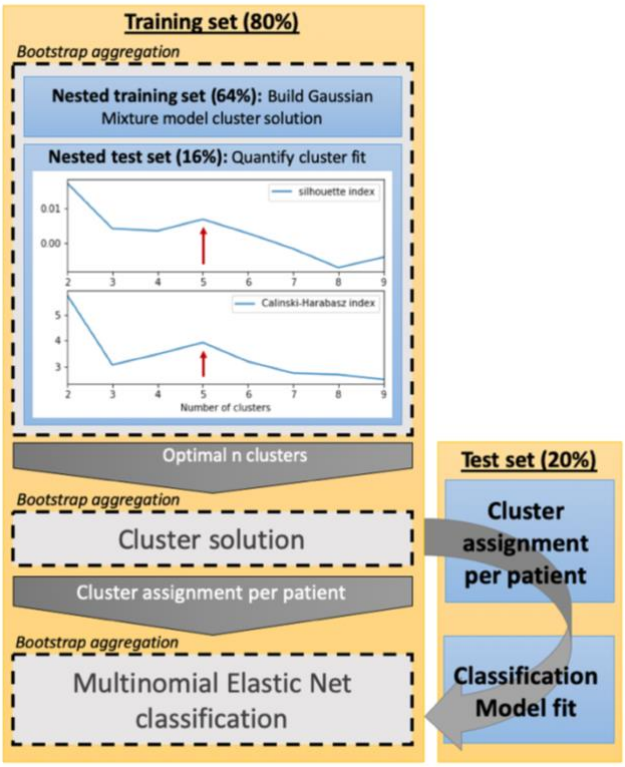

Figure 1. Clustering analysis framework

Method: Using data from N=420 patients ('training sample') diagnosed with major depressive disorder (MDD, n=400) or bipolar disorder (n=20) and assessed at admission and during treatment in a psychiatric hospital we sought to identify a latent cluster structure indicating disease subgroups. 152 features based on Freesurfer segmentations were used to identify a cluster solution using Gaussian Mixture Model (GMM) Clustering and to build a multi-class classifier of cluster membership using Elastic Net regression. Clustering and classification model building were carried out within a 5-fold nested cross-validation (CV) framework and using bootstrap aggregation for stability (see Figure 1).
A four-cluster solution was found to be optimal in all CV folds based on silhouette index. Based on the final GMM models for each CV fold the combination of clusters from each CV fold with the cumulative smallest amount of mean squared error between cluster centroids was identified, and the beta weights from the respective classification models from each fold were averaged to arrive at a consensus model. This aggregated classification model from all CV folds was applied to a second, separate validation set [2] (n=257, 'test sample'). In both samples group differences and correlations between cluster similarity for each patient and Hamilton Depression Rating Scale (HDRS) items responses throughout 6 weeks of treatment were evaluated.

Results: The number of patients most similar to each of the four clusters defined through the aggregated classification model was approximately equal between samples (see Figure 2). Examination of the similarity of each patient to each cluster revealed that there was a strong negative correlation between similarity to cluster 2 and similarity to cluster 4 (r_{train}=.768; r_{test}=.776; see Figure 3). Among the 10 strongest predictors of cluster 2 and cluster 4 membership (see table 1) were cortical thickness in the lateral OFC and entorhinal cortex (cluster 2 < cluster 4), and cortical thickness in the cuneus and paracentral gyrus (cluster 4 < cluster 2). In addition, small left hippocampal volume predicted cluster 2 membership and high cortical thickness in the right parahippocampal gyrus predicted cluster 4 membership.

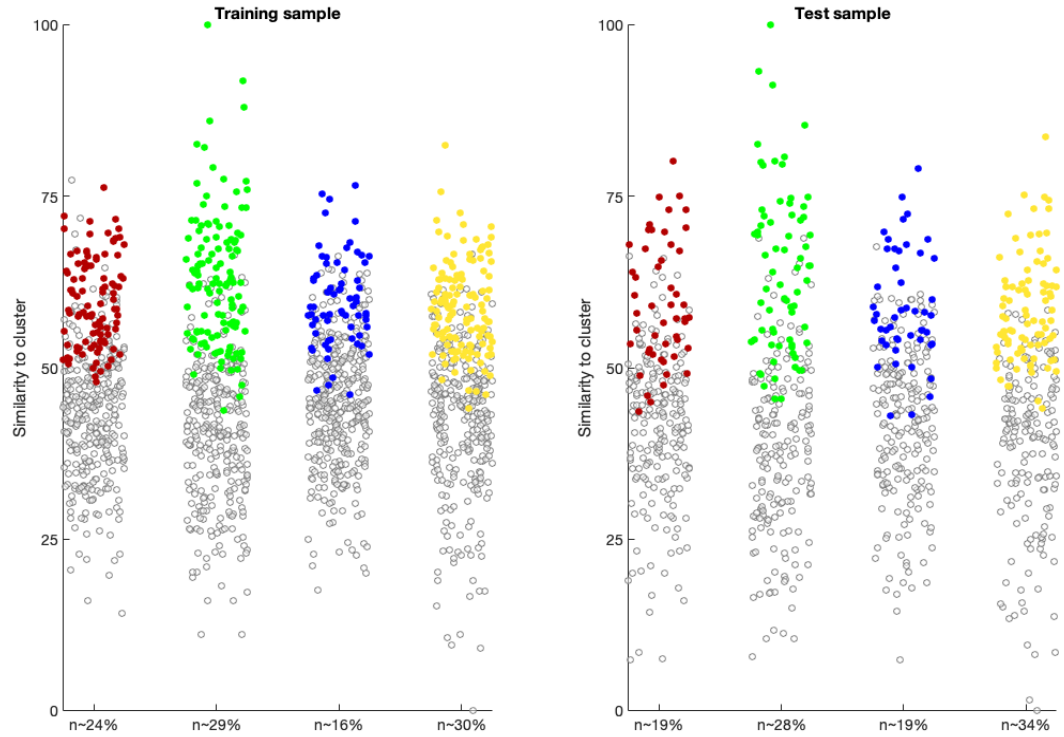

Figure 2 Similarity of all patients to each cluster based on the aggregated classification model. Colored markers represent individuals for whom similarity to this cluster was higher than to the other clusters

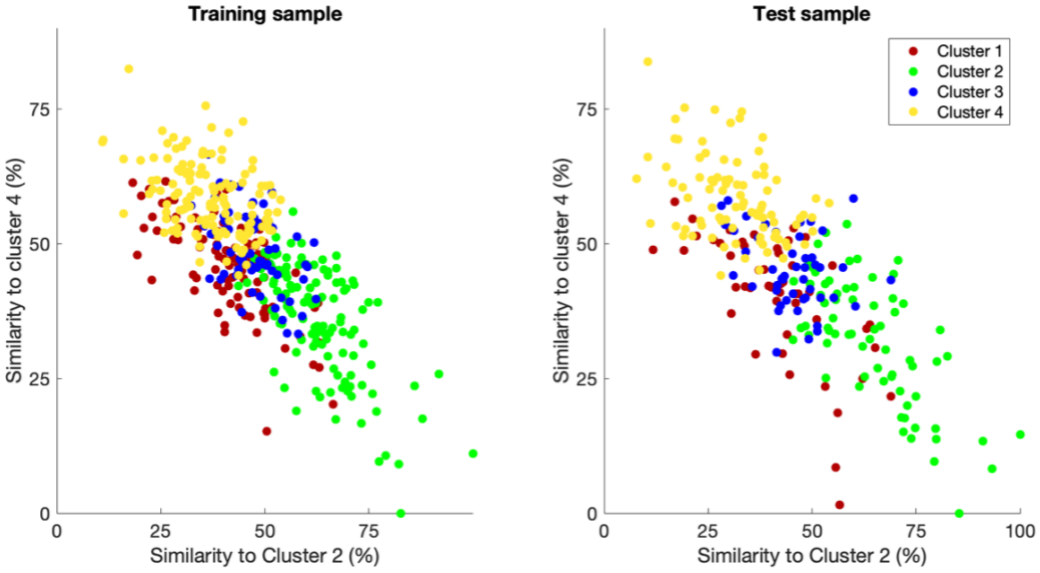

Figure 3 Similarity of all patients to cluster 2 and cluster 4, with colors representing the cluster patients are most similar to.

Table 1 10 Strongest MRI predictors of cluster 2 and cluster 4 membership from the aggregated classification model

Cluster 2	**ß**	**Cluster 4**	**ß**
R_paracentral_thickavg	.0751	L_cuneus_thickavg	-.0682
R_entorhinal_thickavg	-.0774	R_precuneus_thickavg	-.0692
Lhippo	-.0808	L_fusiform_surfavg	.0692
L_parstriangularis_surfavg	.0841	R_lateralorbitofrontal_thickavg	.0726
R_parsopercularis_surfavg	.0849	R_cuneus_surfavg	-.0770
R_rostralanteriorcingulate_thickavg	-.0856	L_entorhinal_thickavg	.0783
R_cuneus_thickavg	.0890	L_transversetemporal_thickavg	-.0842
L_lateralorbitofrontal_thickavg	-.0911	R_parahippocampal_thickavg	.0961
R_isthmuscingulate_surfavg	.1270	L_precuneus_thickavg	-.0976
L_lateralorbitofrontal_surfavg	.1725	L_paracentral_thickavg	-.1099

When dividing patients by whether they were more similar to cluster 2 or cluster 4 higher disease severity was a strong predictor of cluster 2 membership. There were no cluster-by-week interactions for responses to HDRS items in the training sample, but cluster 2 had significantly higher "guilt" and "paranoid" symptoms. In the test sample cluster 2 and cluster 4 also significantly differed in occurrence of "guilt" at week 3, 4, and 5 of treatment, but no significant differences in paranoid symptoms were observed (see Figure 4).

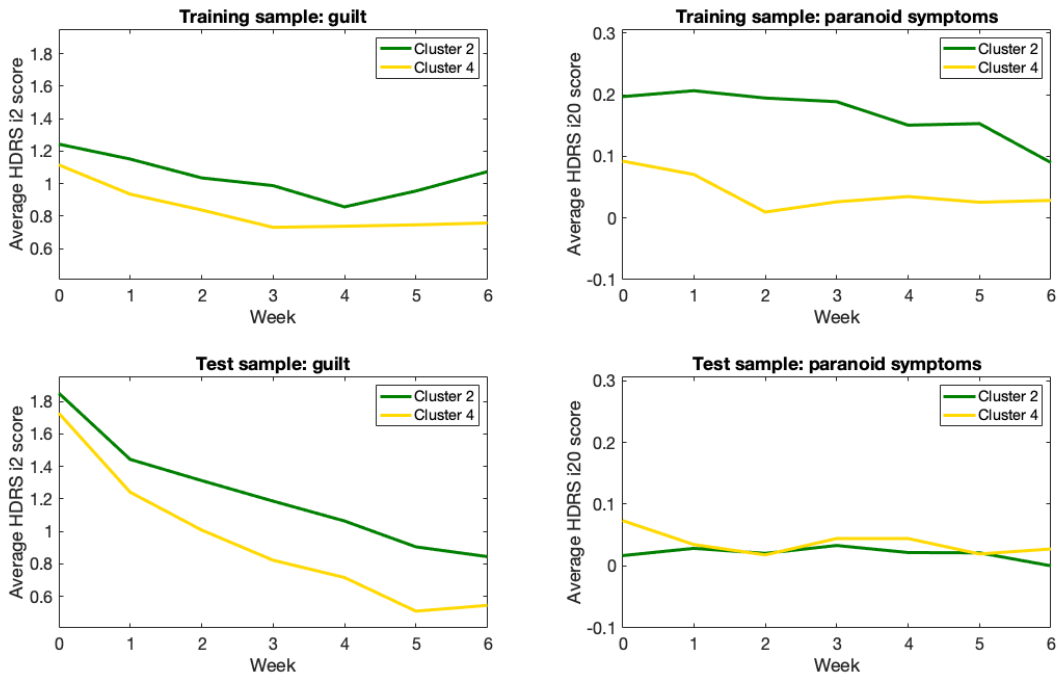

Figure 4 HDRS "guilt" and "paranoid symptoms" over six weeks of treatment for cluster 2 and cluster 4 in the training and test sample

Based on the strong linear relationship observed between cluster 2 similarity and cluster 4 similarity, all patients in the training sample were projected onto a continuous trajectory between cluster 2 and 4 (see Figure 5), and this linear value was used to construct a further linear regression model to be applied to the test set. Elastic Net regression with 5-fold CV and bootstrap aggregation was used.

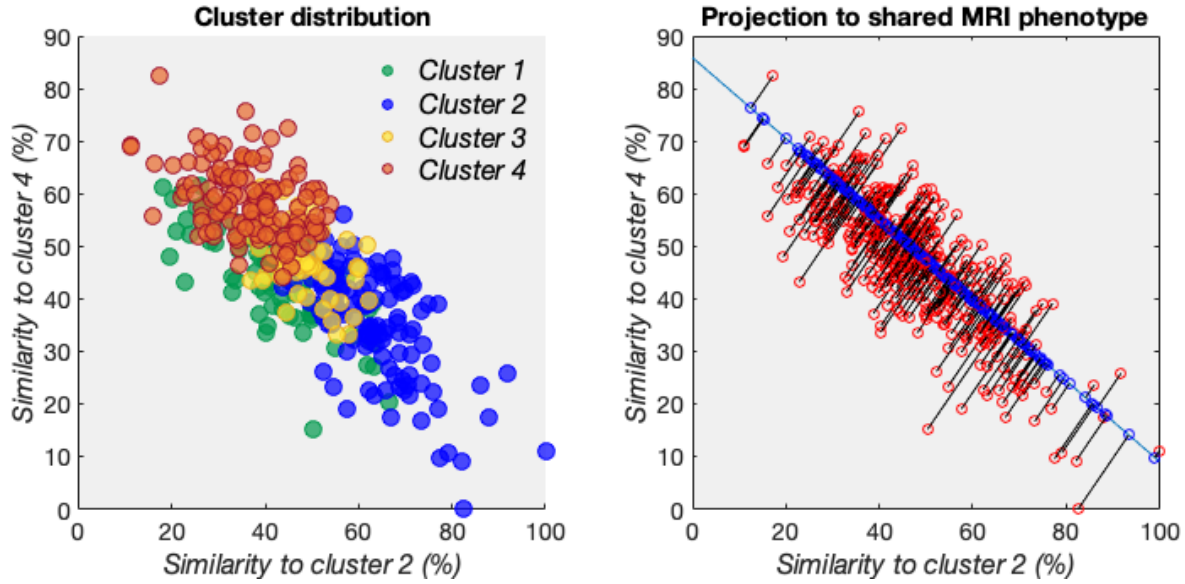

Figure 5 Left: Similarity of all patients in the training sample to cluster 2 and 4. Right: Projection of all patients onto a vector defined by the linear relationship between cluster 2 and cluster 4 similarity

The MRI predictors identified for the continuous MRI phenotype were substantially similar to those observed to differ between cluster 2 and cluster 4 (see Figure 6; table 2). The predicted phenotype correlated at r>.99 with the projected values in the training sample.

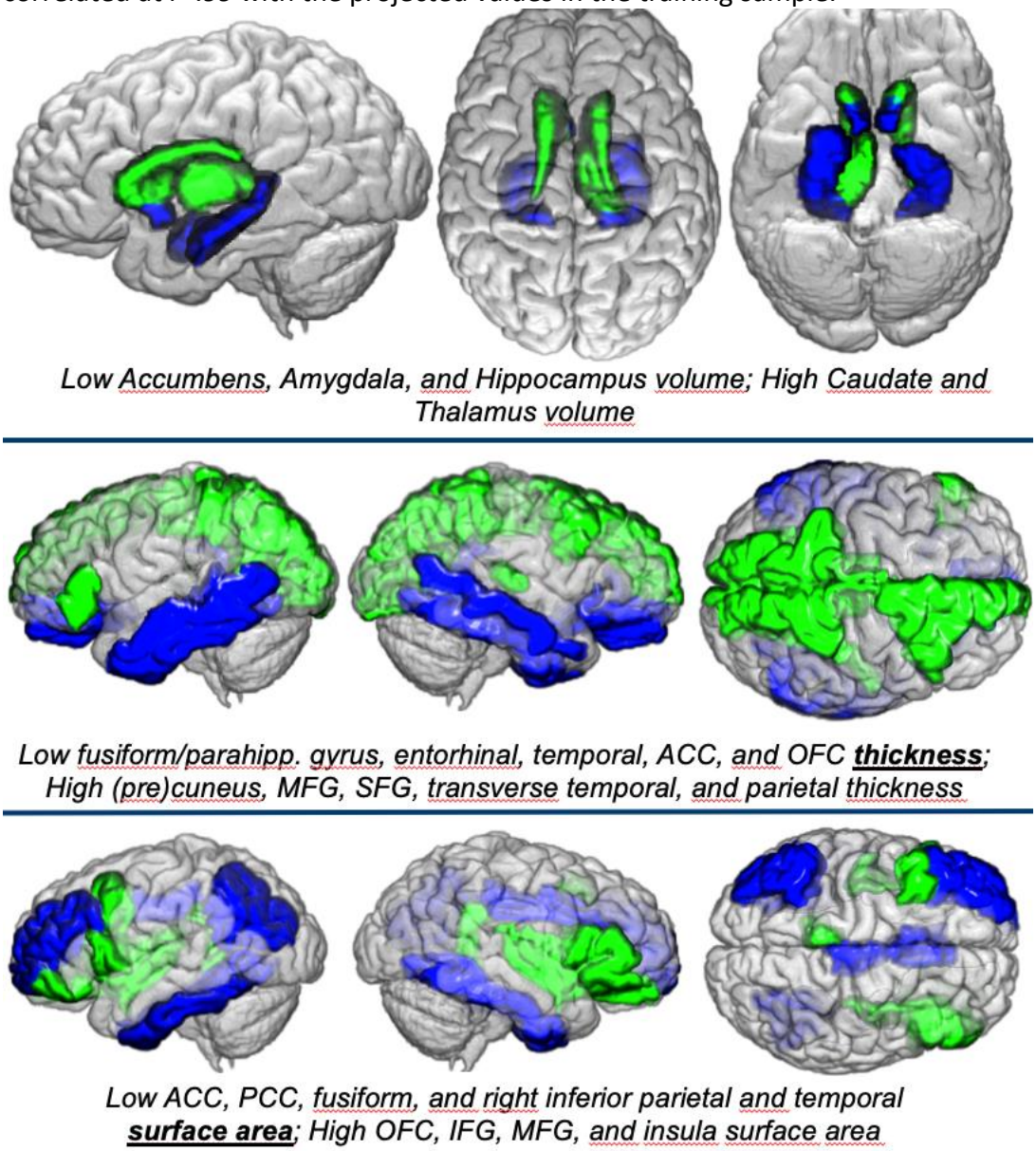

Figure 6 Predictors of the linear MRI phenotype significant based on a null model

Table 2 10 Strongest MRI predictors of the linear projection combining cluster 2 and cluster 4 similarity

Cluster 2	**ß**
L_lateralorbitofrontal_surfavg	1.6369
R_isthmuscingulate_surfavg	1.5981
L_parstriangularis_surfavg	1.3781
R_cuneus_thickavg	1.3665
R_superiorparietal_thickavg	1.3347
L_transversetemporal_thickavg	1.3219
Lhippo	-1.3080
R_parsopercularis_surfavg	1.3072
R_fusiform_surfavg	-1.2957
L_paracentral_thickavg	1.2383

When using only patients' responses to the Hamilton Depression Rating Scale items 1 to 21 to predict linear MRI phenotype values the predicted phenotype correlated at r=.1293 with the projected phenotype values. The strongest predictors in this model were items 2 (feelings of guilt) and 3 (suicide) (see Figure 7).

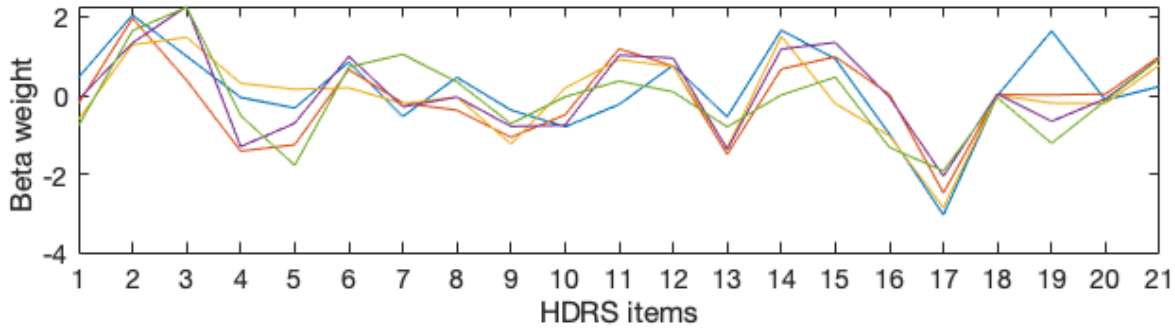

Figure 7 Beta weights for the prediction of the linear MRI phenotype using only HDRS item responses in 5 cross-validation folds

Exploratory follow-up analyses applying the MRI model of the linear phenotype to the test sample revealed that the linear phenotype correlated with HDRS item 2 and 3 responses between 4 and 6 weeks after begin of inpatient treatment (baseline) (see figure 8).

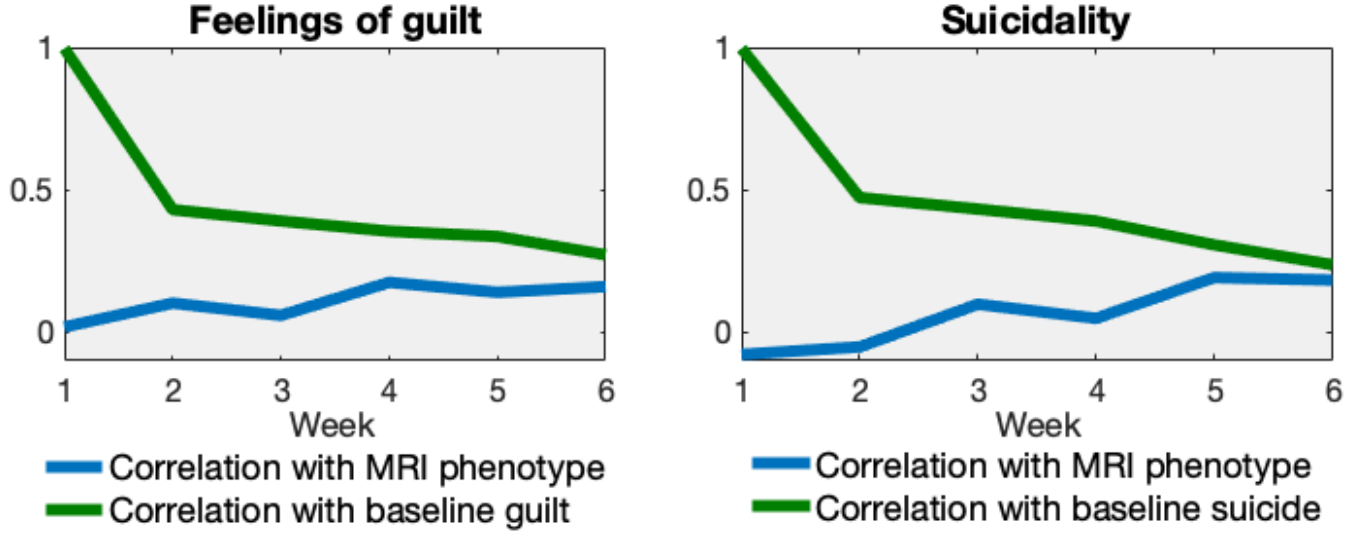

Figure 8 Correlation of the linear MRI phenotype and of baseline values for HDRS items 2 and 3 with follow-up values during 6 weeks of inpatient treatment

Conclusion: Using cortical thickness, cortical surface area, and subcortical volume measures a set of clusters describing a structural MRI phenotype with clinical relevance in two separate samples of mood disorder patients was identified. The two clusters described here were defined based on measures including hippocampus volume, and showed differences in development of guilt throughout six weeks of treatment. Differences in Paranoid symptoms were also observed in the training sample, which were not seen in the test sample. This was likely due to the exclusion of individuals with psychotic symptoms in the second sample. A linear phenotype defined based on two of the four clusters was associated with follow-up symptoms in the test sample. This phenotype may show utility in identifying mood disorder patients at increased risk for treatment resistance and failure to show remission.

[1] Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., ... & Aboyans, V. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1211-1259.

[2] Dunlop, B. W., Binder, E. B., Cubells, J. F., Goodman, M. M., Kelley, M. E., Kinkead, B., ... & Pace, T. W. (2012). Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. Trials, 13(1), 106